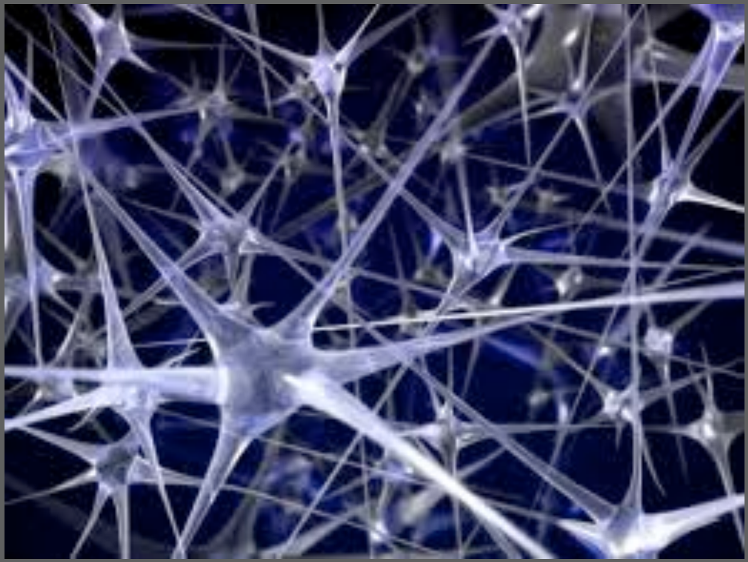
Complex Systems Tools for Describing Spatio-temporal Changes of Microbial Community Dynamics in Freshwater Environments



Donna Rizzo
School of Engineering
University of Vermont

Roadmap

Methods

- Microbial community fingerprinting
- Geostatistical & nonparametric statistical methods

Goal: Develop new computational tools that improve:

- Spatial interpretation of water resources problems
- Understanding of human-induced changes on natural systems and the way we make decisions about natural resources

Complex Systems Tools (artificial neural networks)

- Classification
- Clustering
- Leverage spatial and temporal auto-correlation in multiple data type

Decision-making Applications

- **Characterizing and monitoring landfill leachate fringes**

  Mouser, P.J., D.M. Rizzo, G. Druschel, S.E. Morales, P. O'Grady, N.J. Hayden and L. Stevens, (2010) *Water Resources Research*.

  Pearce, A.R., D.M. Rizzo and P.J. Mouser, (2011) *Water Resources Research*.

- **Feature detection (I.D. conditions necessary for onset of algal blooms)**

  Pearce, A.R., M.C. Watzin, G. Druschel, L. Stevens and D.M. Rizzo, (in Review) "Identifying conditions associated with cyanobacteria blooms in Missisquoi Bay, Lake Champlain, USA, using a modified Self-Organizing Map" *Limnology and Oceanography.*

- **Linking stream geomorphology, habitat and biological assessments**

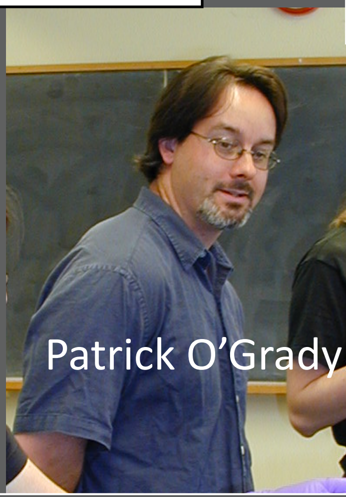
  Mathon, B.R., G.F. Pinder, L. Stevens, M. Kline, G. Alexander and D.M. Rizzo, (In Review), "Classifying Vermont stream habitat condition using a generalized regression neural network", *Journal of the American Water Resources Association.*

THE UNIVERSITY of VERMONT

Vermont EPSCoR

Department of Civil &

Environmental Engineering



Patrick O'Grady

Donna Rizzo Paula Mouser

Cassella Waste Services

Schuyler Landfill, N.Y. Bernie Nadeau

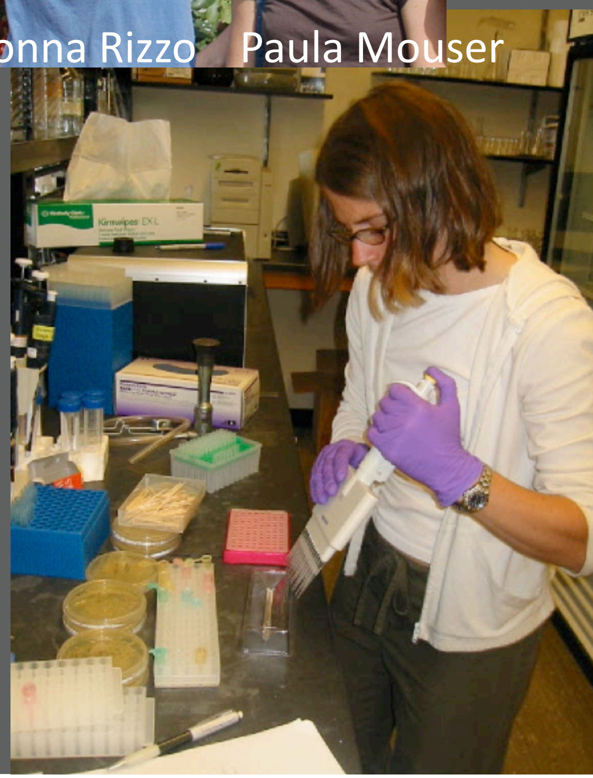
New York State DEC

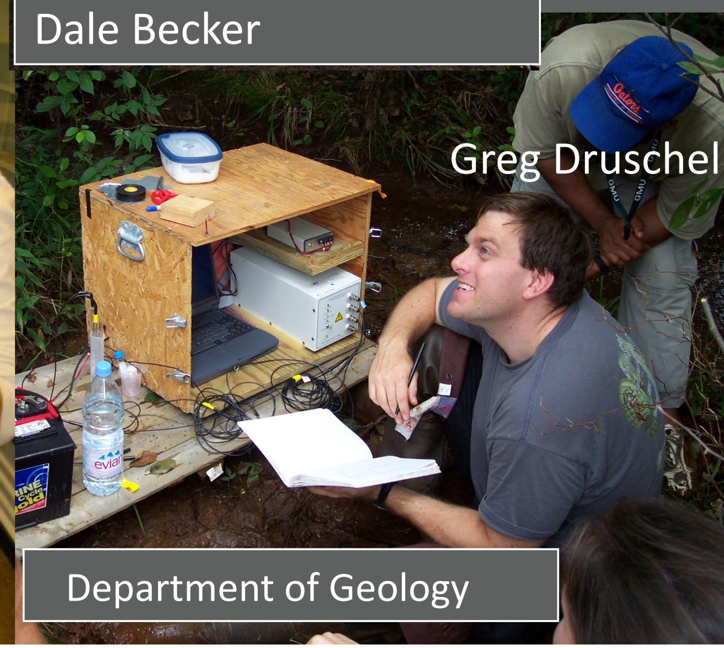
Dale Becker

Department of Biology

Lori Stevens

Brooke Schwartz





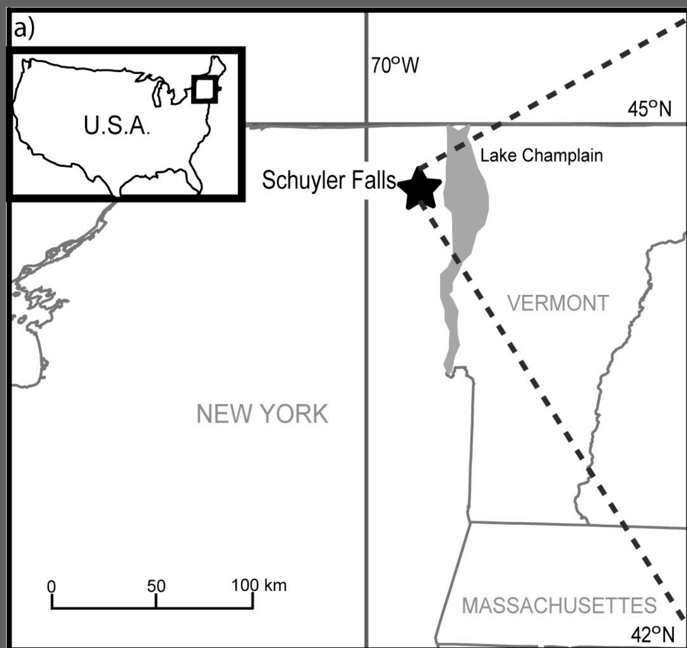
Greg Druschel

Department of Geology

Research Motivation

Monitoring at multi-contaminant sites is expensive and incomplete.

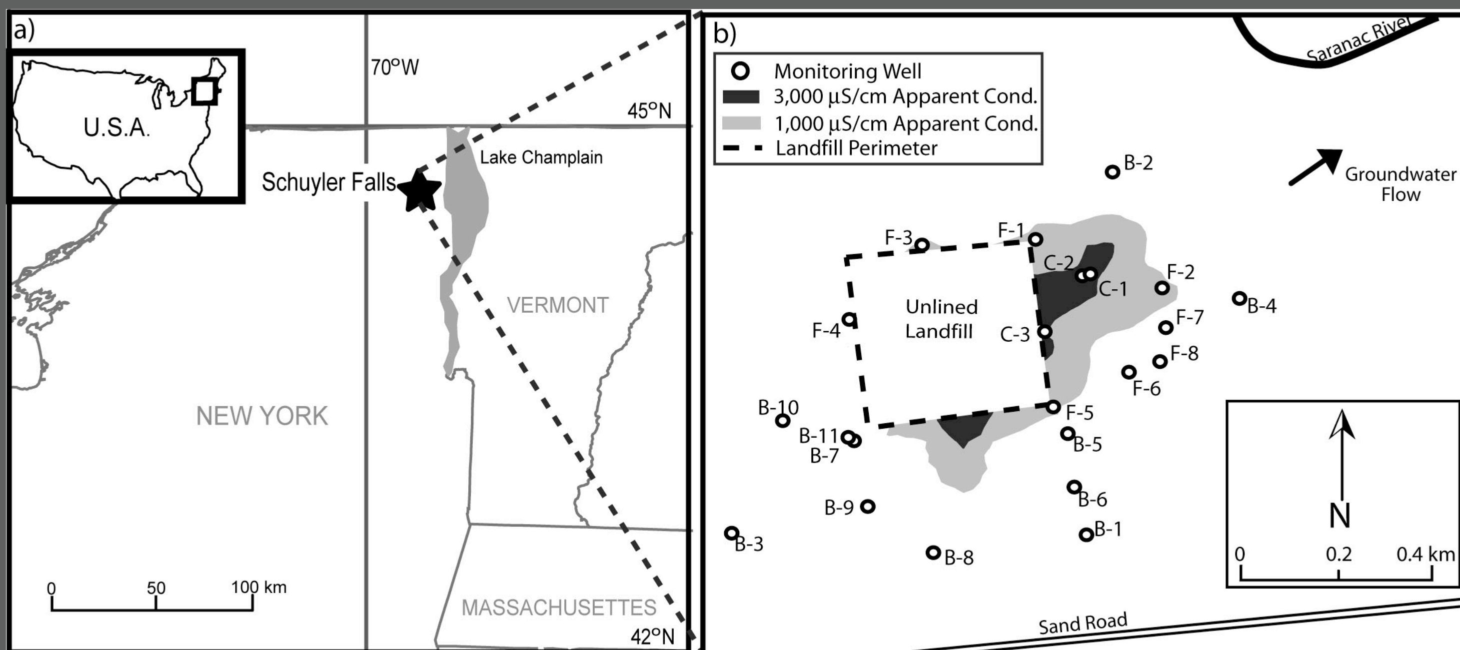
- What should you monitor at landfills?
- How to describe overall extent of contamination?
- Geo-spatial interpolation is not possible.



Motivation for the research

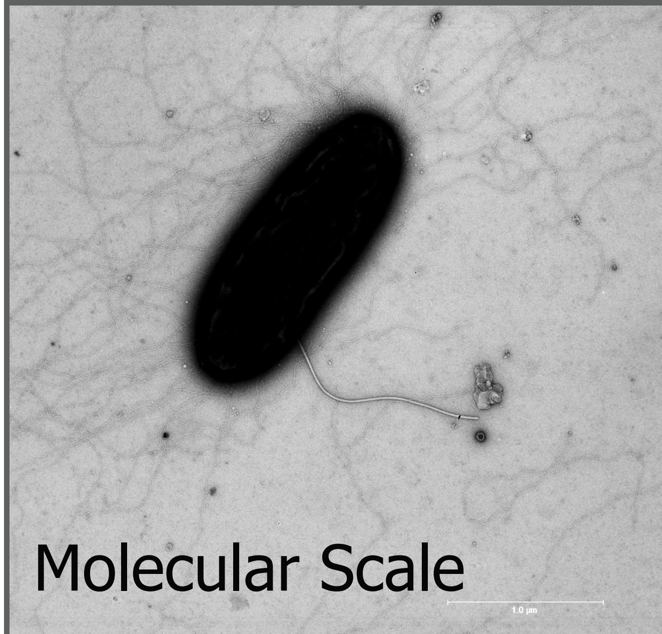
Monitoring at multi-contaminant sites is expensive and incomplete.

- What should you monitor at landfills?
- How to describe overall extent of contamination?
- Geo-spatial interpolation is not possible.



Idea: Use microbial information to develop a more systematic, and effective method for characterizing and monitoring aquifers with multiple contaminants.

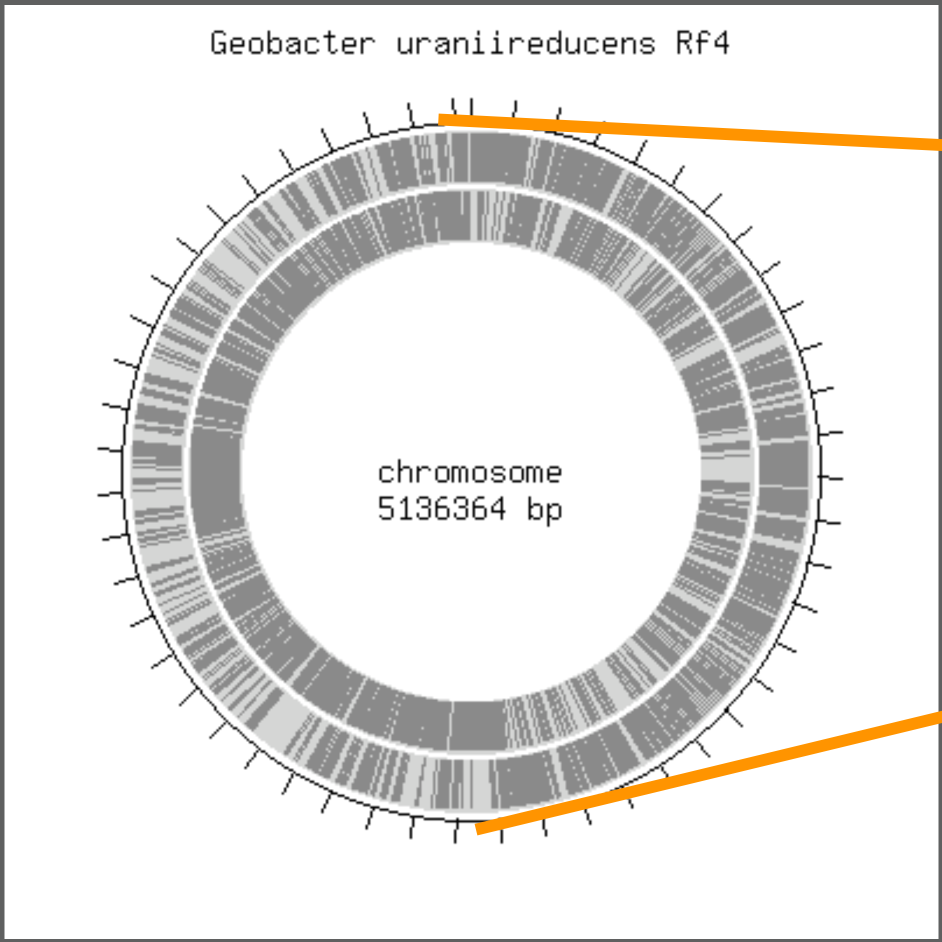
The "Undefined" Intermediate



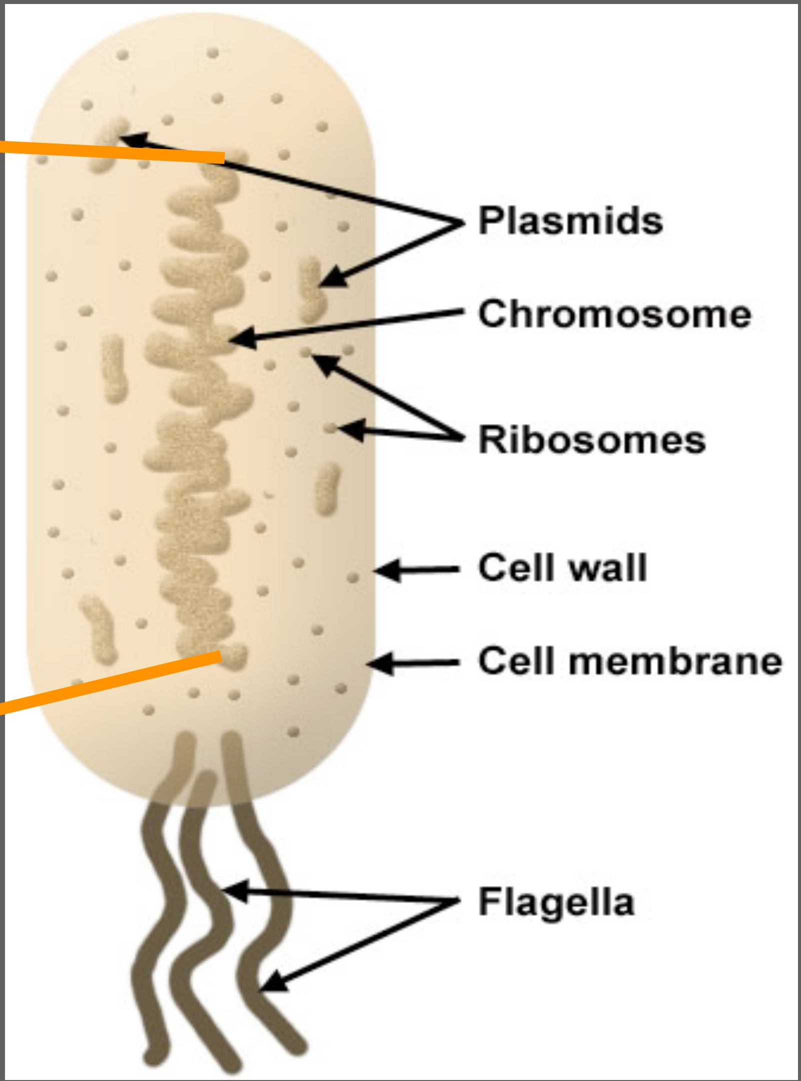
- Species Capable of Contaminant Transformations
- Molecular-level Understanding of Ecology and Physiology
- Metabolic and Constraint-based Genomic Models

- Do microbial community profiles enhance our ability to detect/predict contamination?
- How are they correlated to hydrogeochemisty?
- Are they useful for parameter estimation and predictive models?

Methods: Molecular Genetic-Based Techniques

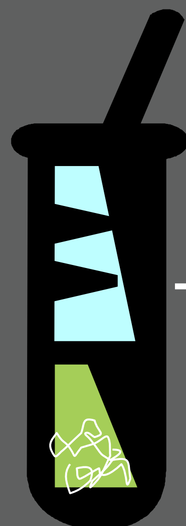


The genome of G. *uraniireducens*
http://gib.genes.nig.ac.jp/img/crcl/Gura_RF4/

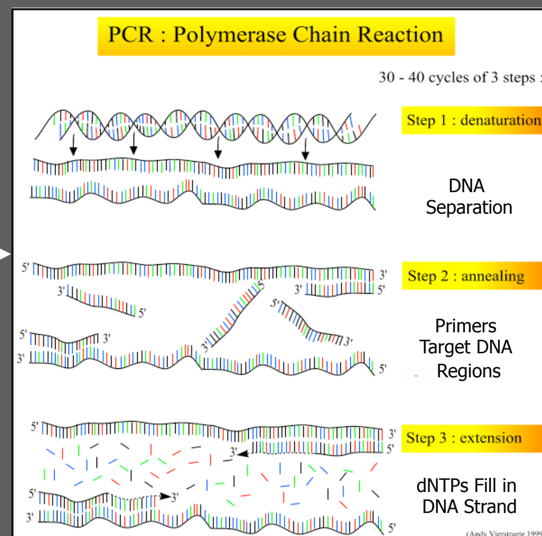


Applied to Environmental Engineering

Groundwater/Soil Sampling

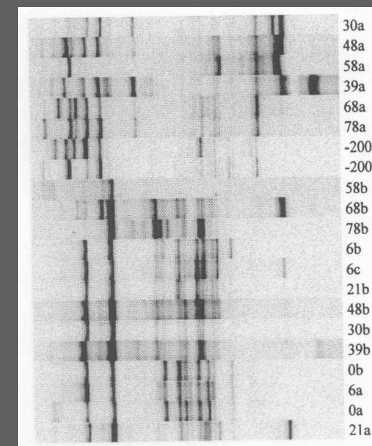


Nucleic Acid Extraction & Purification

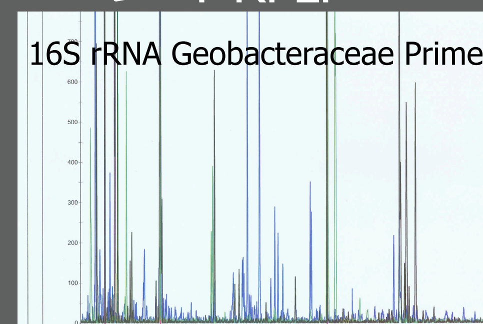


DNA/RNA Amplification Using PCR

DGGE



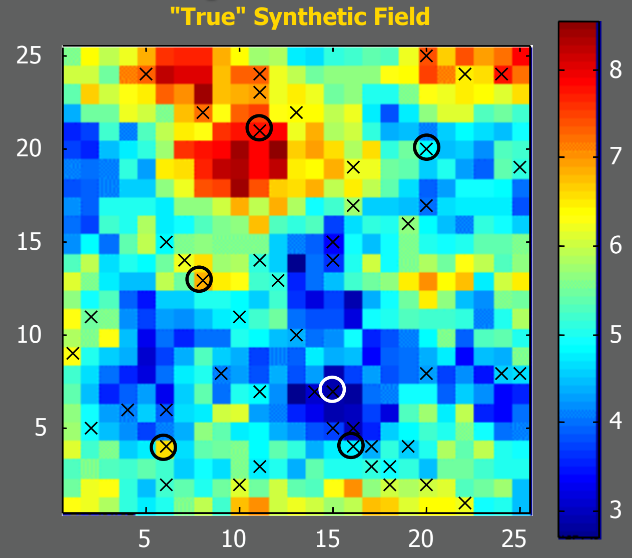
T-RFLP



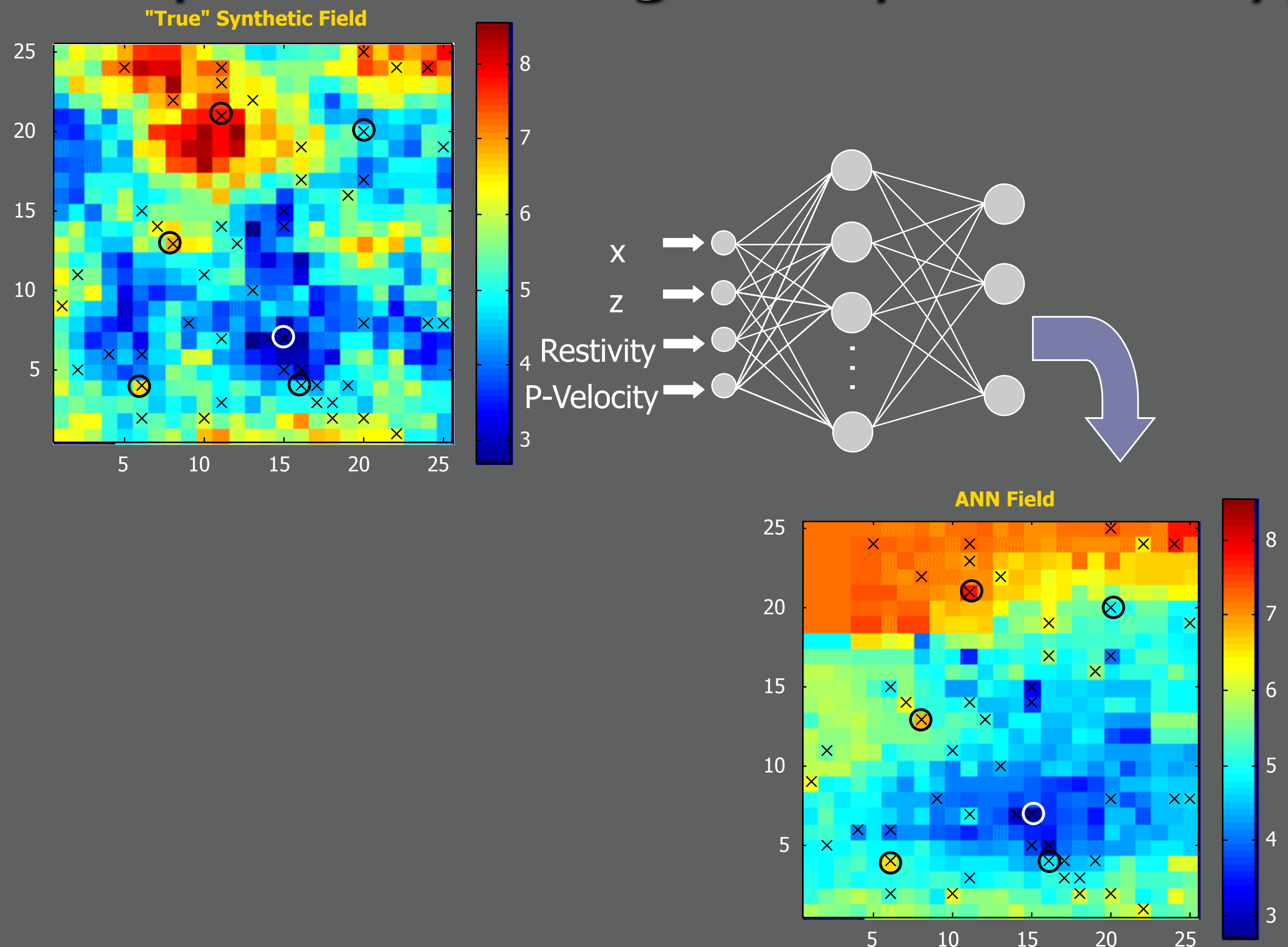
Visualization-Quantification

Röling, W.F.M., van Breukelen, B.M., Braster, B.M., Lin, B., Verseveld, H.W. (2001). *Applied & Env. Microb.* 67(10), 4619-4629.

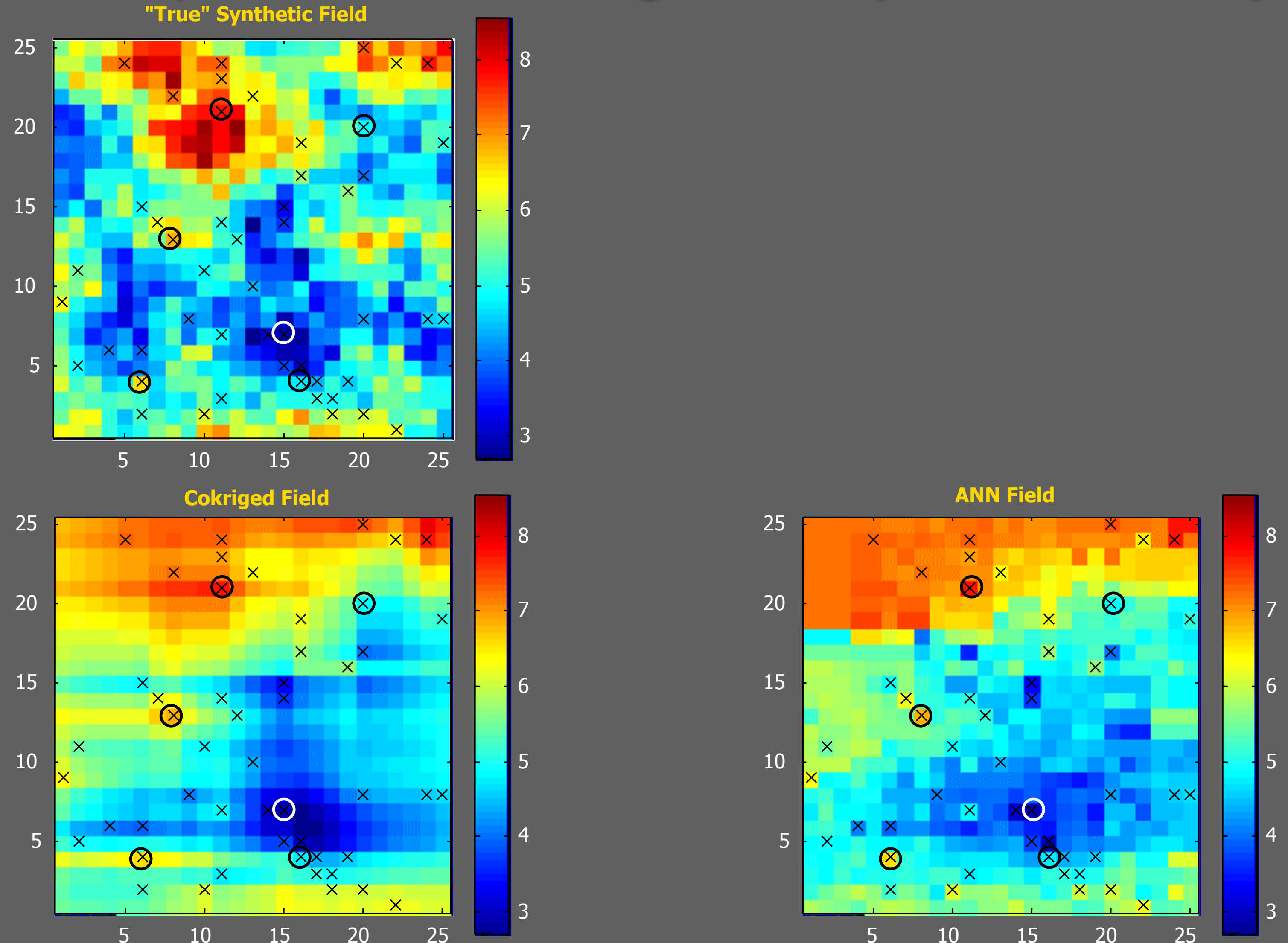
Develop New Estimation Tool capable of using Multiple Data Types



Develop New Estimation Tool capable of using Multiple Data Types



Develop New Estimation Tool capable of using Multiple Data Types



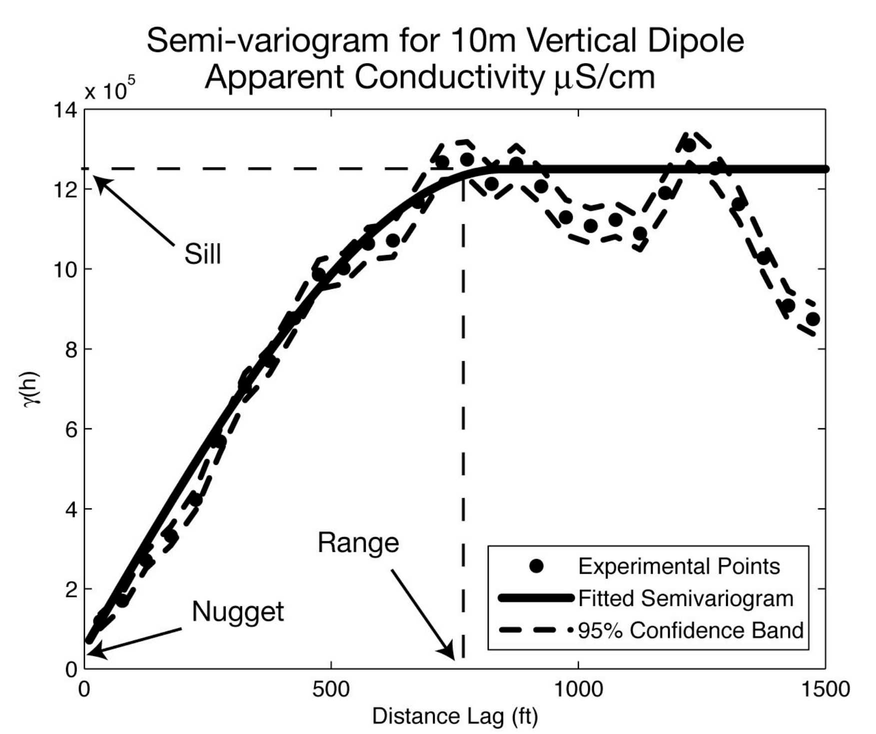
Spatial Structure and Estimation

- Develop New Computational Tools (Replicate Co-Kriging)
  - Estimate Parameter and
  - Error Variance at unsampled locations

- Calculate Semi-Variance
  - Estimate Range

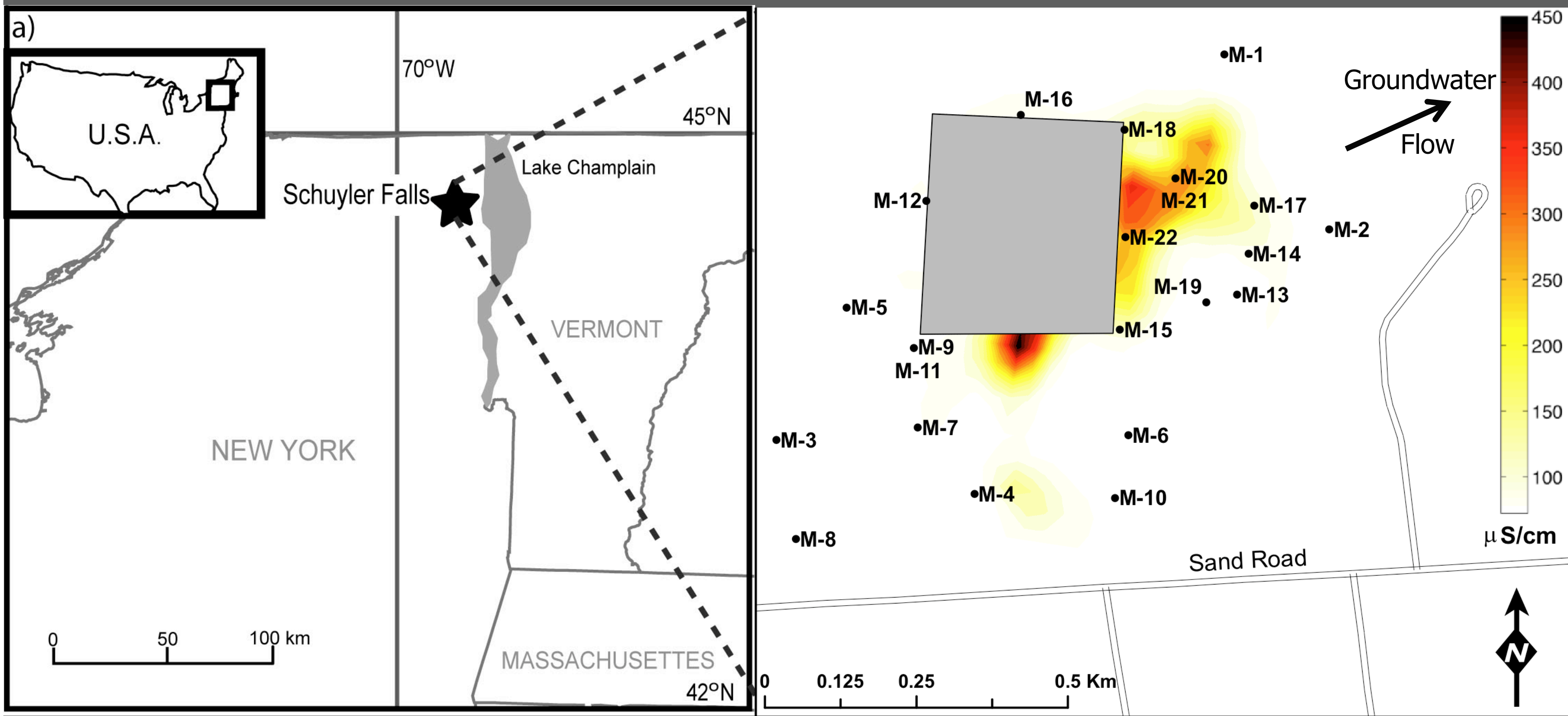
    Distance where samples are no longer correlated
  - Estimate Sill

    Variance where samples are no longer correlated
  - Fit Model



Applications

Schuyler Falls Landfill, NY



Parameters Measured

Groundwater Microbial Community

16S rRNA
Bacteria
Archaea
Geobacteraceae

T-RFLP Analysis
Sequencing

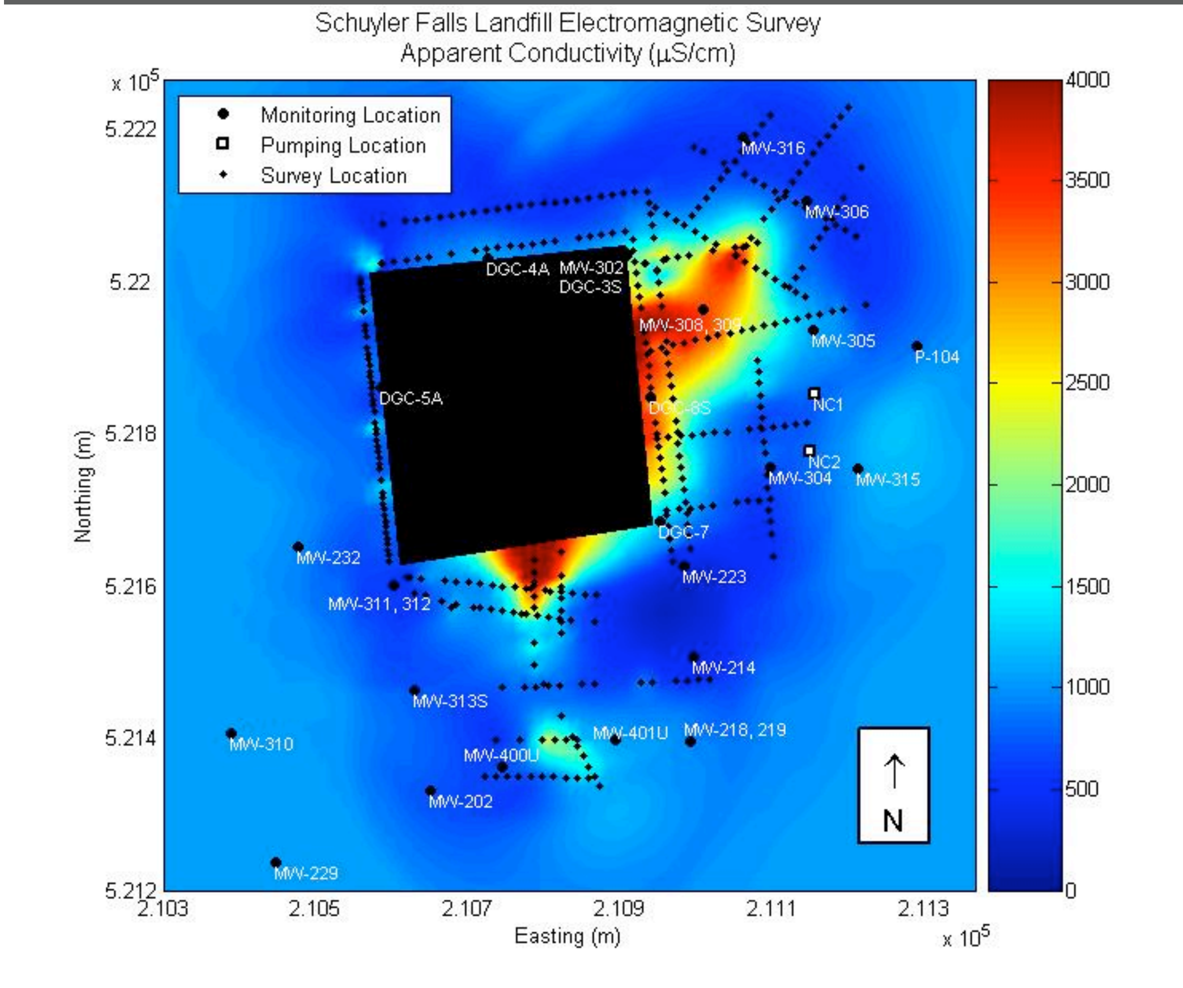
Groundwater Hydrogeochemistry

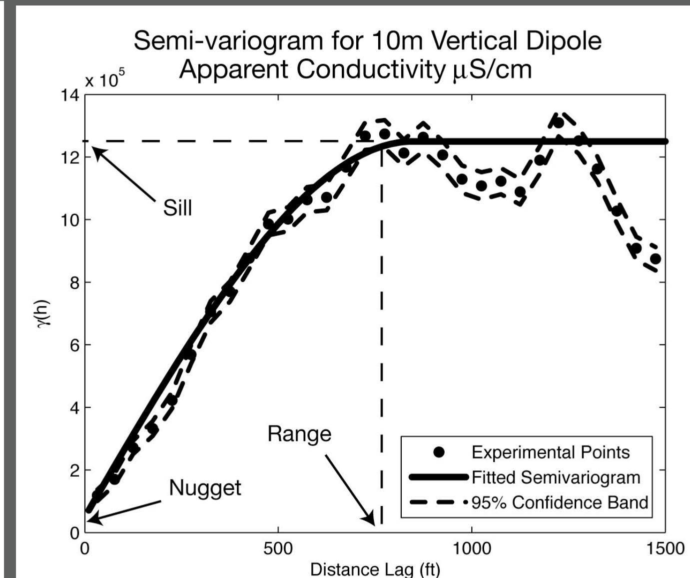
Temperature
pH
Turbidity
Alkalinity
TDS
Conductivity
Cl
COD
BOD_5

Al
Ca
Fe
K
Mg
Mn
Na
Si

TOC
Eh
NO_3^-
SO_4^{2-}
NH_3
NH_4^+
TKN
Phenols
VOCs

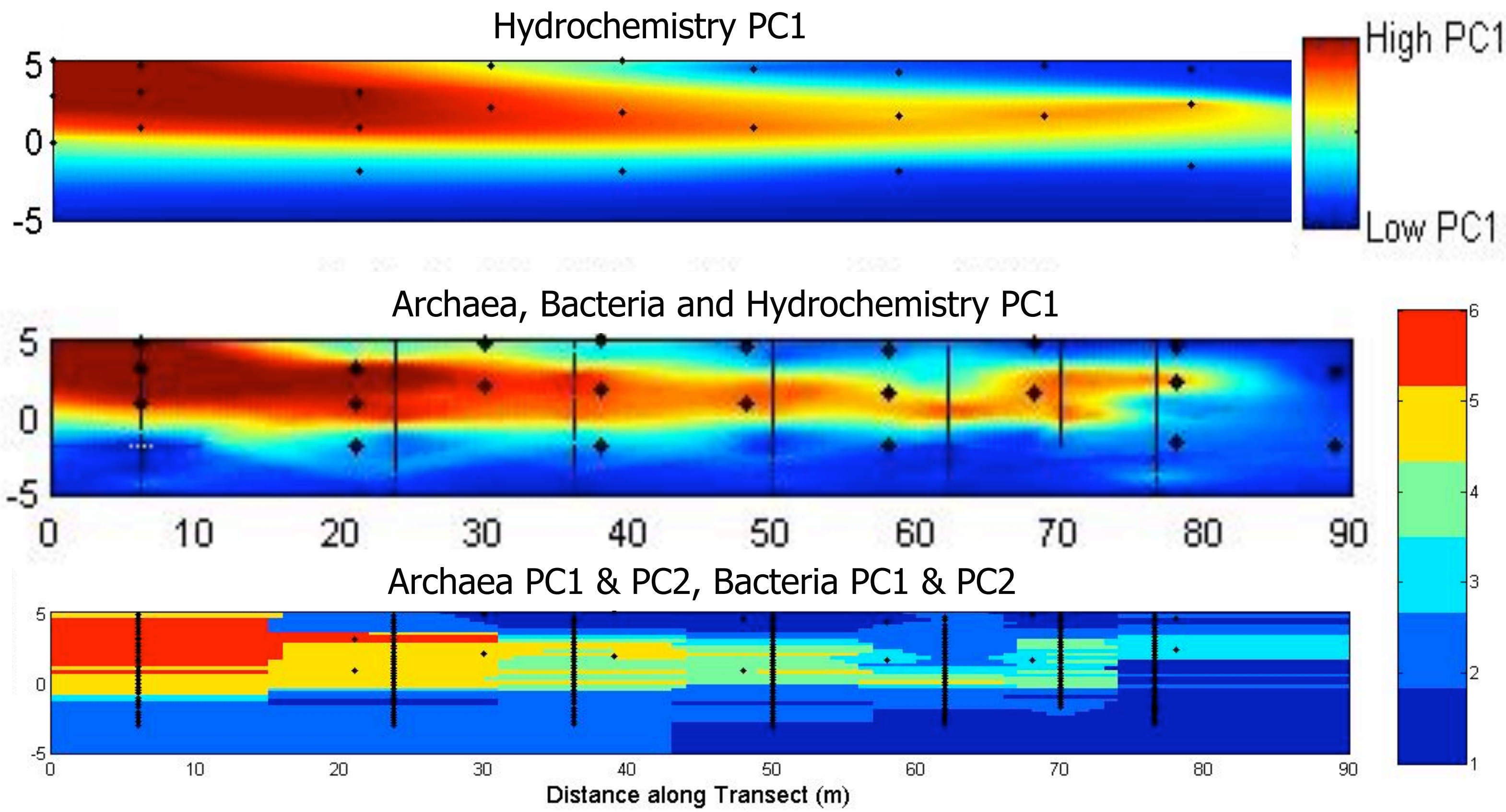
Geophysical Site Characterization





- 664 EM-34 Survey Points (10/20 m horizontal & vertical dipole)
- Used to create detailed map of "contamination"

Applications



Mouser, P.J., Rizzo, D.M., Röling, W.F.M., van Breukelen, B.M., (2005). *Environmental Science & Technology,* 39, 7551-7559.

Besaw, L.E. and D.M. Rizzo, (2007). *Water Resources Research*, 43, W11409, DOI: 10.1029/2006WR005509.

Subsurface characterization of groundwater contaminated by landfill leachate using microbial community profile data and a non-parametric decision-making process

Andrea R. Pearce[1], Donna M. Rizzo[1] and Paula J. Mouser[2]

Photo Credit: http://www.resourcesystemsconsulting.com/blog/

(1) School of Engineering, University of Vermont
(2) Civil and Environmental Engineering and Geodetic Science, The Ohio State University

How does one groundwater characterize contamination at landfills?

	Low	Medium	High	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18	M19	M20	M21	M22
Spec. Cond., μS/cm [2]	< 1000	1000-5000	> 5000	L	L	L	L	L	L	L	L	L	L	M	L	L	L	M	M	M	M	M	H	H	H
NH_3 [3]	<1	1-100	>100	L	L	L	L	L	L	L	L	L	L	L	M	M	M	M	L	M	M	M	H	H	H
Alkalinity as $CaCO_3$ [2]	< 500	500-1200	> 1200	L	L	L	L	L	L	L	L	L	L	M	M	L	L	M	M	M	M	M	H	H	H
ORP, mV [3]	> 50	+50 to -50	<- 50	L	M	L	M	L	H	M	M	M	L	L	L	M	H	M	M	M	M	H	M	H	M
Fe [3]	< 10 [4]		> 10	L	L	L	L	L	L	L	L	L	L	L	H	H	H	H	H	H	H	L	H	H	L
Total Phenols [2]	<= 0.005 [4]		> 0.005	L	L	L	L	L	L	L	L	L	L	L	L	L	H	H	L	L	L	L	H	H	H
BOD [2]	<= 2 [4]	2 - 20	>= 20	L	L	L	L	L	L	L	L	L	L	L	M	M	H	M	H	M	H	M	H	H	H
COD [2]	< 10	10 - 100	> 100	L	L	L	L	L	L	M	M	L	M	L	M	M	M	M	M	M	M	M	H	H	H
Mouser et al. [2010] Classification				B	B	B	B	B	B	B	B	B	B	B	F	F	F	F	F	F	F	F	C	C	C

2 - Indicates a significant difference was observed between B and F, C locations ($p < 0.05$ Tukey-Kramer test for multiple comparisons among means)

3 - Indicates a significant difference was observed between B,F and C locations ($p < 0.05$ Tukey-Kramer test for multiple comparisons among means)

4 - Detection limit

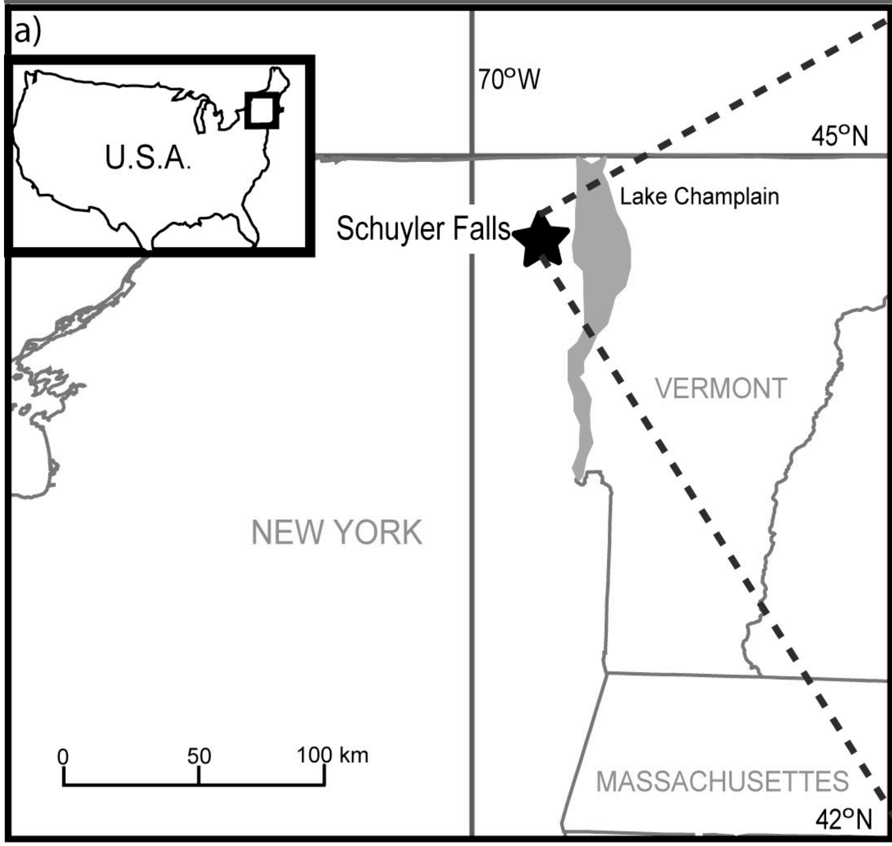
Application

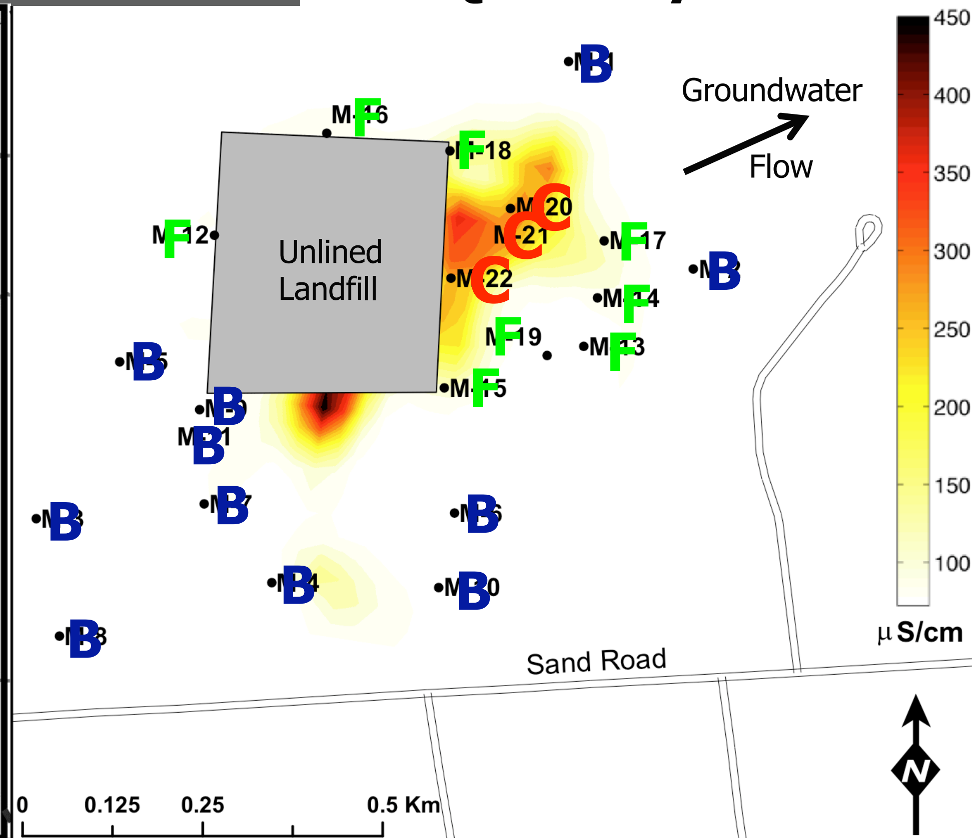
3 - Impacted

8 - Fringe

11 - Background

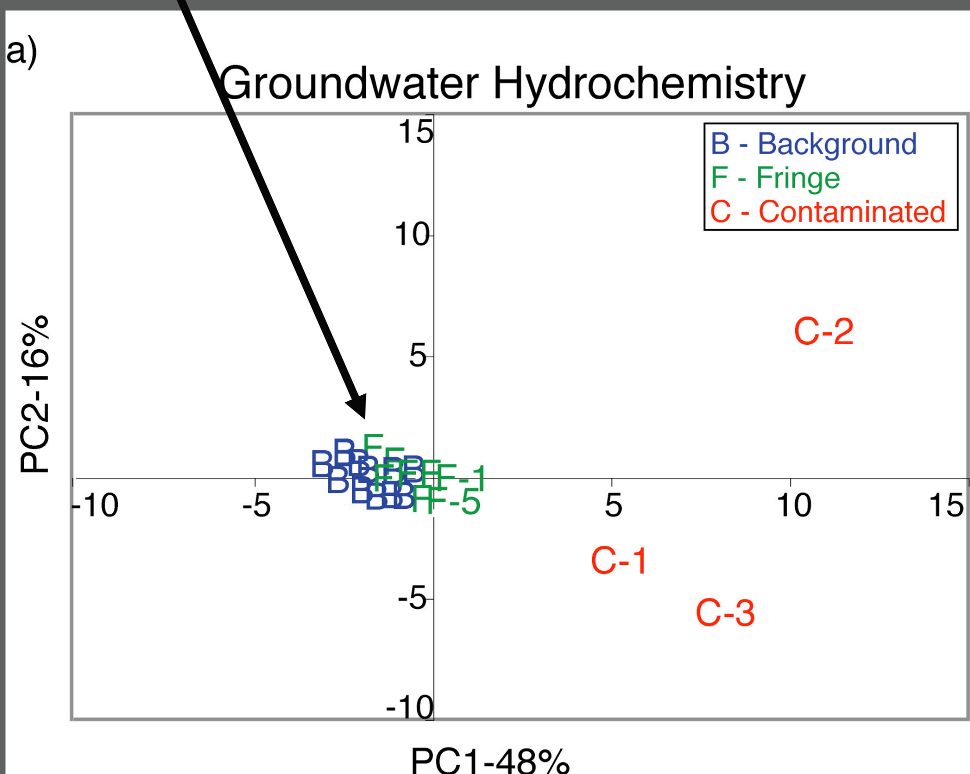
22 Monitoring Locations
5 Quarterly Events





Can we Detect "Fringe" Contamination with Hydrogeochemistry Information?

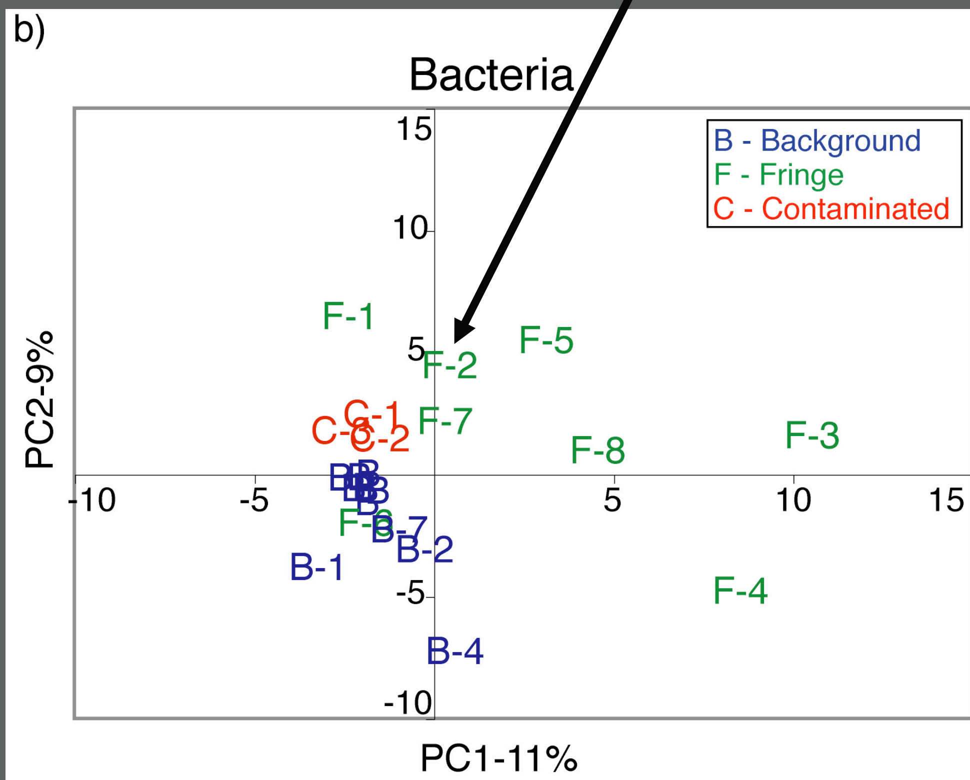
NO-
Fringe & Background Statistically Similar



- PC1 & PC2 separate contaminated samples
- PC1 correlations
  - TDS, Conductivity, Cl, Mg, Hardness, Alkalinity COD, TOC, NH_3
- PC2 correlations
  - Organic-N, Phenols

Can we Detect "Fringe" Contamination with Bacterial Profiles?

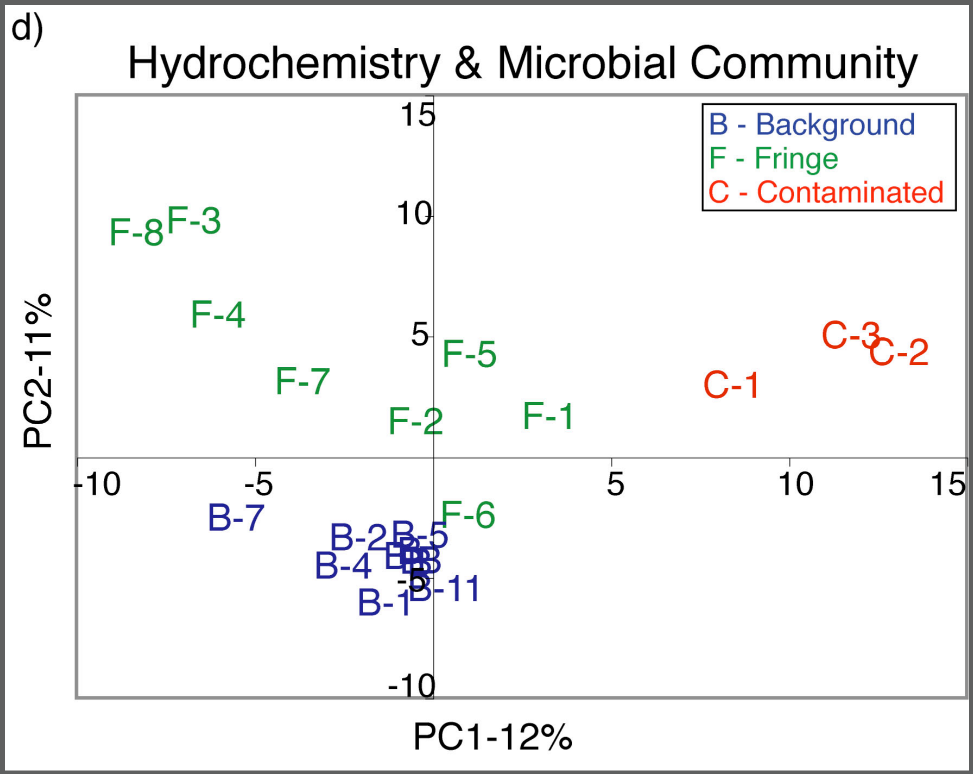
YES -
Fringe Locations Significantly Different



- Fringe Separate across PC1 and PC2, some overlap
- PC1 correlations
  - Correlated T-RFs:
    B160, B178, B279, B439, B450
- PC2 correlations
  - Correlated T-RFs (Positive)
    B167, B172, B203, B430, B504
  - Negative
    B64, B103, B106, B125, B428

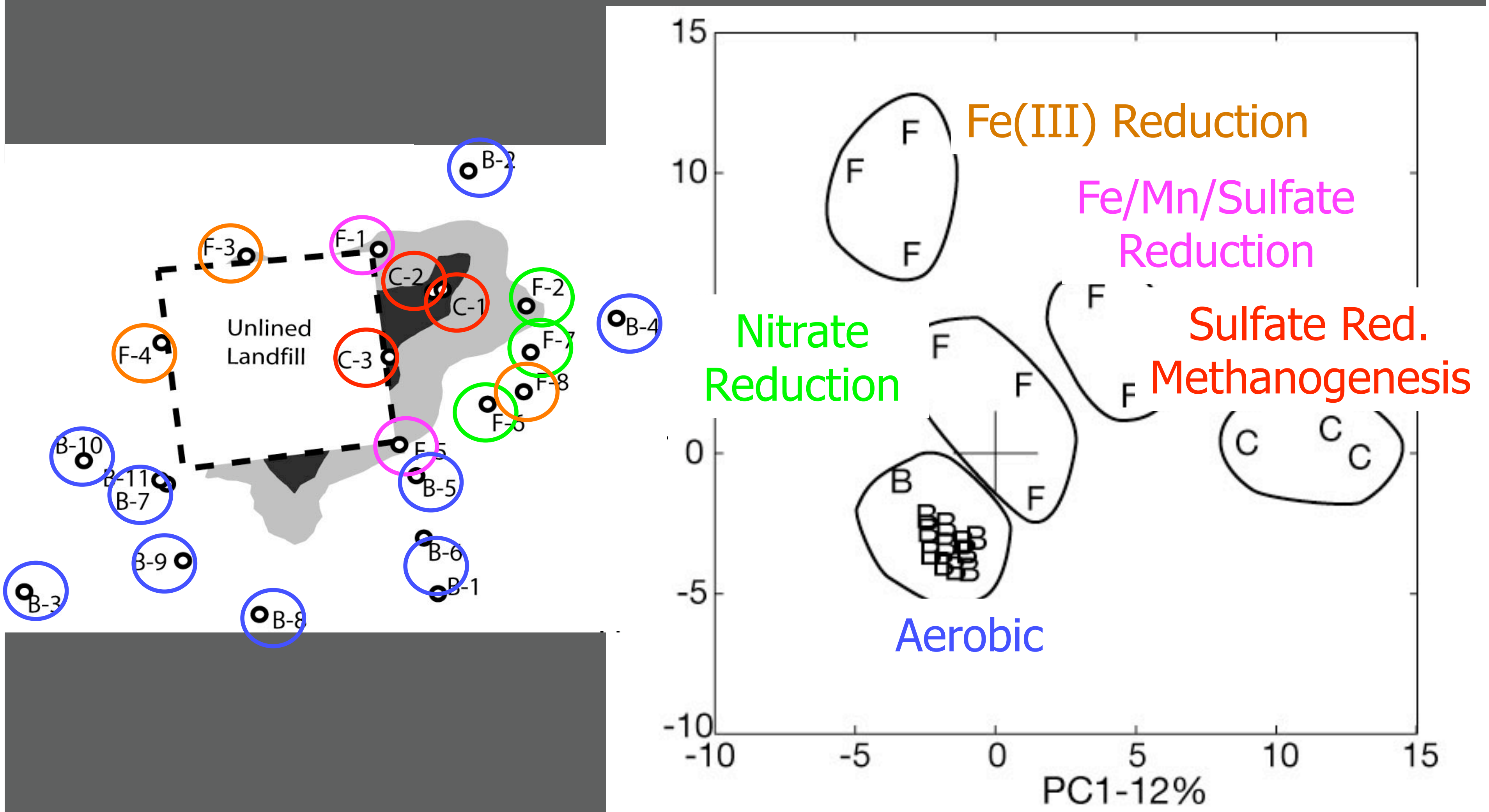
Does Adding Microbial Data and Hydrochemistry Enhance Detection?

YES-
All 3 Groups Statistically Different

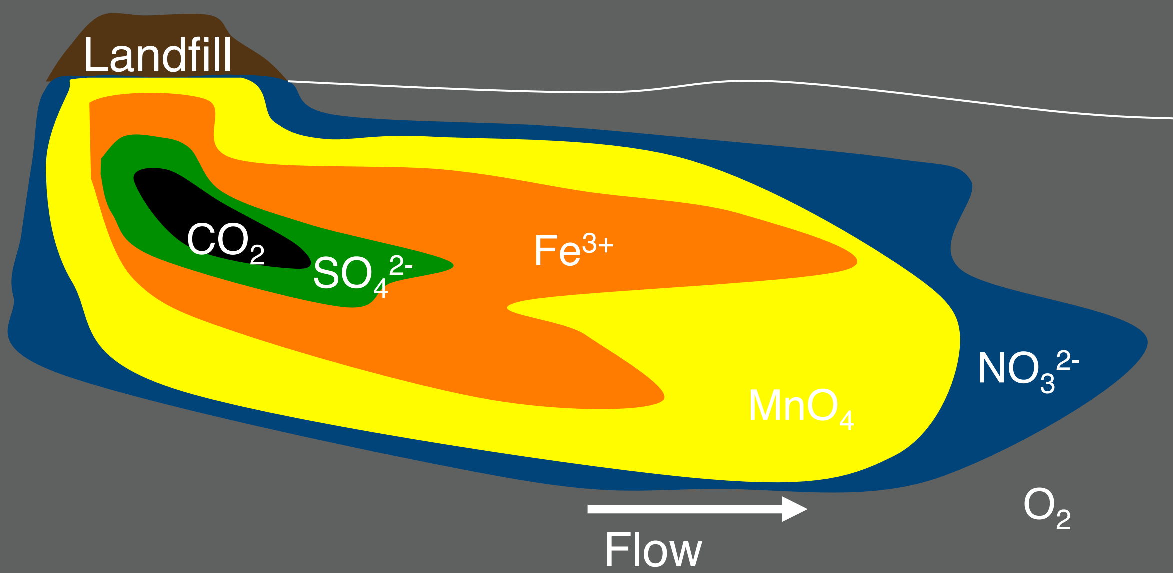


- PC1 separates contaminated
  - Leachate Indicators, Eh, Mn, SO_4^{2-}
  - Correlated T-RFs: G505, B244, B122, G80, B168, G165, A244
- PC2 separates fringe
  - Fe, NO_3^-, pH
  - Correlated T-RFs: B121, B160, G424, A118, G510, A144, B492, G484, B279, B470

Combined Data Enhance Detection and Characterization of Pollution



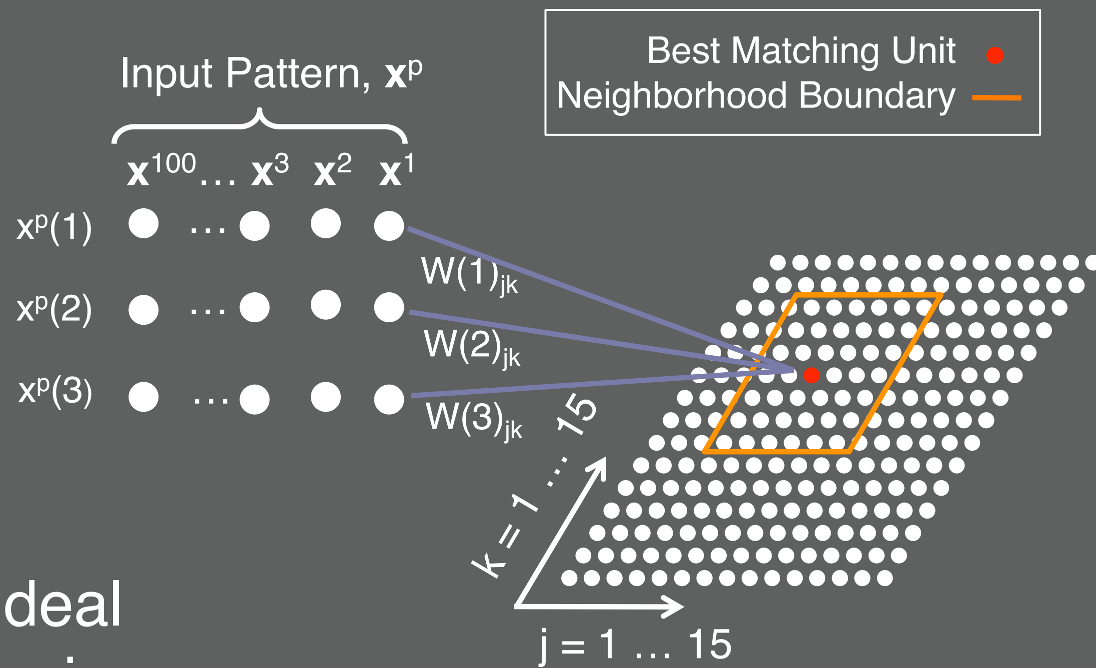
Objective: To *cluster* groundwater samples with similar microbial community profiles to compliment traditional hydrochemical analyses for the purpose of delineating spatial zones of groundwater contamination.



Few methods exist for incorporating these data at the field scale (due, in part, to the high dimensionality and complex relationships with physiochemical parameters).

The Self-Organizing Map (SOM) is a Clustering Artificial Neural Network

Why Clustering ?



•SOM architecture has a cognitive basis and is ideal for exploratory data analysis

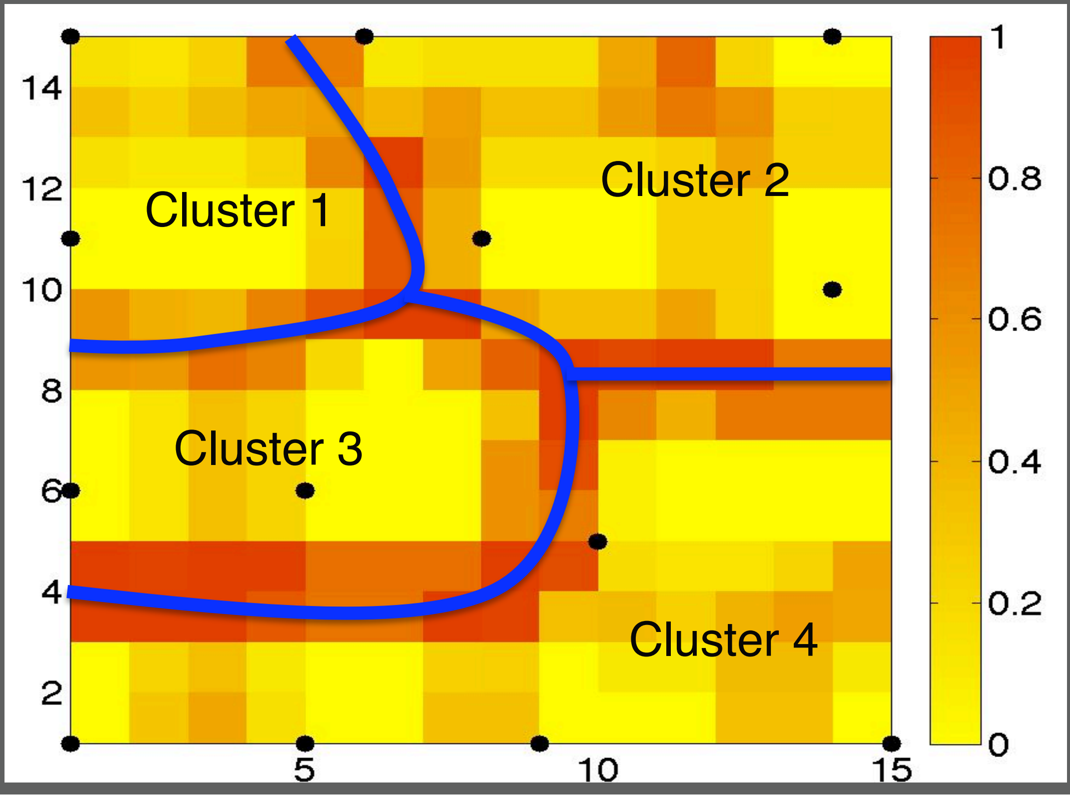
•Do not need to know the # of groups or group attributes prior to analysis

•Few, if any statistical assumptions

Methods

How many clusters are optimal?

- Depends on your research question?
- Many methods exist for comparing groupings (most rely on some ratio of the between : within group variability



Methods

How many clusters are optimal?

- ⌘ Depends on your research question?
- ⌘ Many methods exist for comparing groupings (most rely on some ratio of the between : within group variability

Found a nonparametric MANOVA –

- ✸Accepts groups with different numbers of members (unbalanced designs)
- ✸Can use any distance metric
- ✸Can have more variables than samples
- ✸p based on permutation
- ✸Method not available in commercial statistical packages

Anderson, M. (2001), A new method for non-parametric multivariate analysis of variance, *Austral Ecology, 26(1), 32-46.*

McArdle, B., and M. Anderson (2001), Fitting Multivariate Models to Community Data: A Comment on Distance-Based Redundancy Analysis, *Ecology, 82(1), 290-297.*

				M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18	M19	M20	M21	M22
Spearman Rank PCA	Hierarchical	2	4.58	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2
		3	5.26	1	1	1	1	1	1	1	1	1	1	1	3	3	3	2	3	3	3	2	2	2	2
		4	5.62	1	1	1	1	1			1	1	1		3	3	3	2	3	3	3	2	2	2	2
	K-Means	2	5.15	1	1	2	1	2	2	2	1	2	1	2	1	1	1	2	1	1	1	2	2	2	2
		3	5.86[1]	3	3	1	3	1	1	1	3	1	3	1	3	3	3	2	3	3	3	2	2	2	2
		4	5.57	1	1	1	1	1			1		1		3	3	3	2	3	3	3	2	2	2	2
	SOM	2	4.73	2	2	1	2	1	1	1	2	1	2	1	2	2	2	2	2	2	2	2	2	2	2
		3	5.86[1]	3	3	1	3	1	1	1	3	1	3	1	3	3	3	2	3	3	3	2	2	2	2
		4	5.60	1	1	1	1	1		1	1	1	1		3	3	3	2	3	3	3	2	2	2	2
	Weighted SOM	2	8.08	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2
		3	10.47	1	1	1	1	1	1	1	1	1	1	1	3	3	3	2	3	3	3	2	2	2	2
		4	9.85	1	1	1	1	1			1		1		3	3	3	2	3	3	3	2	2	2	2

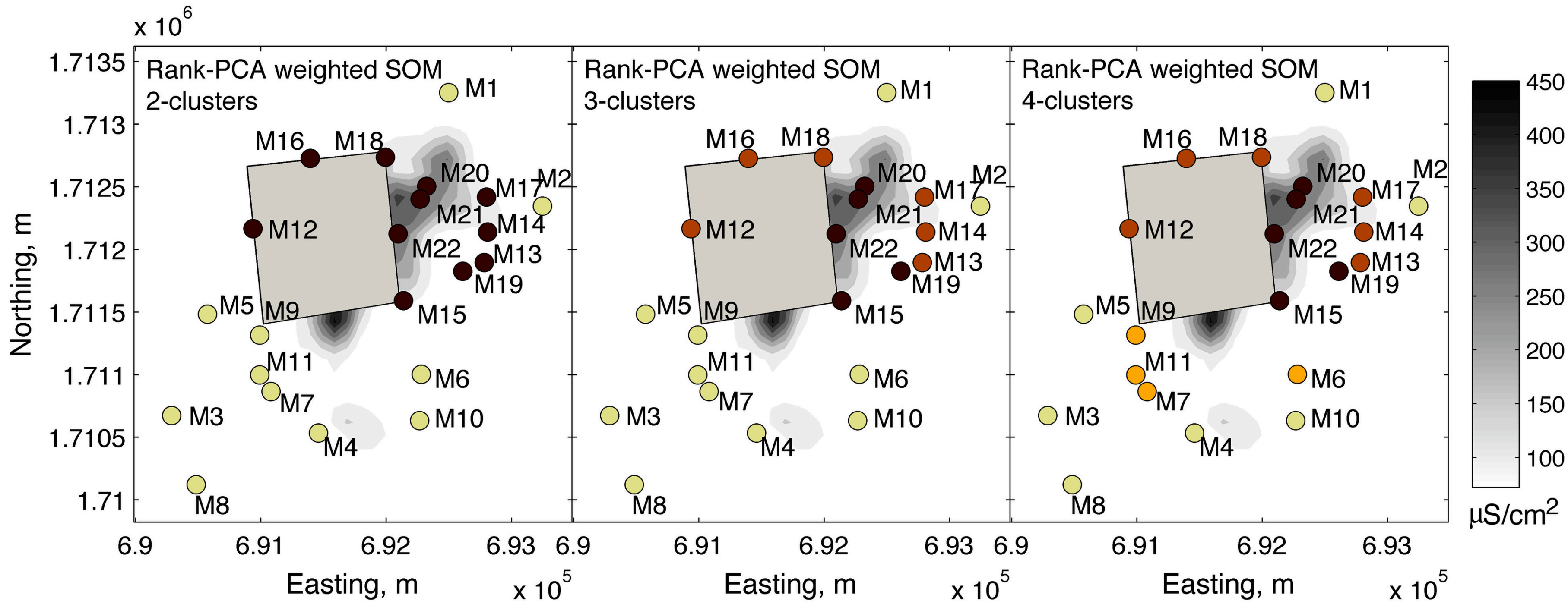
1 - F above 95% confidence limit among comparable samples

	Low	Medium	High	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18	M19	M20	M21	M22
Spec. Cond., μS/cm [2]	< 1000	1000-5000	> 5000	L	L	L	L	L	L	L	L	L	L	M	L	L	L	M	M	M	M	M	H	H	H
NH_3 [3]	<1	1-100	>100	L	L	L	L	L	L	L	L	L	L	L	M	M	M	M	L	M	M	M	H	H	H
Alkalinity as $CaCO_3$ [2]	< 500	500-1200	> 1200	L	L	L	L	L	L	L	L	L	L	M	M	L	L	M	M	M	M	M	H	H	H
ORP, mV [3]	> 50	+50 to -50	<- 50	L	M	L	M	L	H	M	M	M	L	L	L	M	H	M	M	M	M	H	M	H	M
Fe [3]	< 10 [4]		> 10	L	L	L	L	L	L	L	L	L	L	L	H	H	H	H	H	H	H	L	H	H	L
Total Phenols [2]	<= 0.005 [4]		> 0.005	L	L	L	L	L	L	L	L	L	L	L	L	L	H	H	L	L	L	L	H	H	H
BOD [2]	<= 2 [4]	2 - 20	>= 20	L	L	L	L	L	L	L	L	L	L	L	M	M	H	M	H	M	H	M	H	H	H
COD [2]	< 10	10 - 100	> 100	L	L	L	L	L	L	M	M	L	M	L	M	M	M	M	M	M	M	M	H	H	H
Mouser et al. [2010] Classification				B	B	B	B	B	B	B	B	B	B	B	F	F	F	F	F	F	F	F	C	C	C

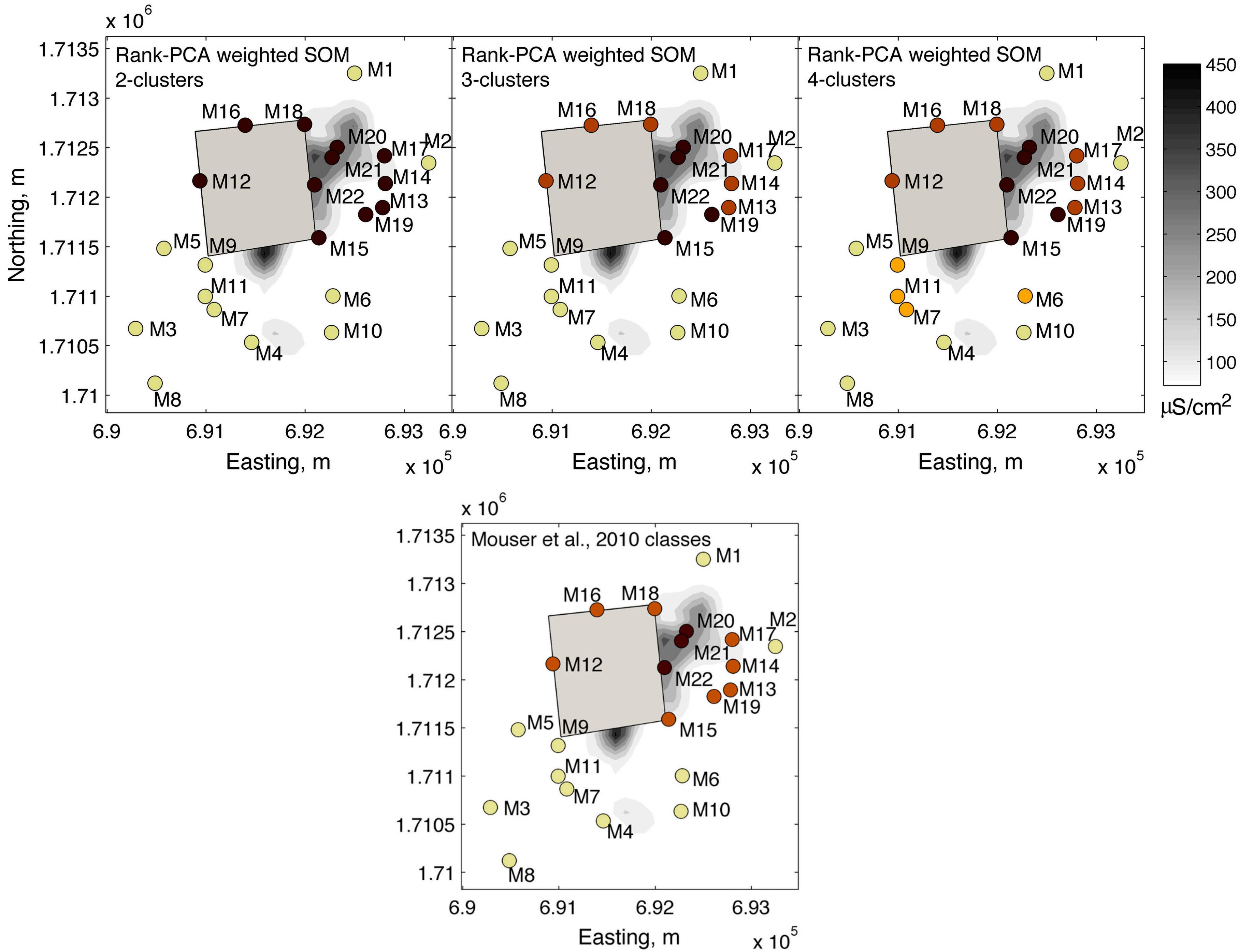
2 - Indicates a significant difference was observed between B and F, C locations (p < 0.05 Tukey-Kramer test for multiple comparisons among means)

3 - Indicates a significant difference was observed between B,F and C locations (p < 0.05 Tukey-Kramer test for multiple comparisons among means)

4 - Detection limit



x 10^6
Rank-PCA weighted SOM
2-clusters
Rank-PCA weighted SOM
3-clusters
Rank-PCA weighted SOM
4-clusters
M1
M2
M3
M4
M5
M6
M7
M8
M9
M10
M11
M12
M13
M14
M15
M16
M17
M18
M19
M20
M21
M22
1.7135
1.713
1.7125
1.712
1.7115
1.711
1.7105
1.71
Northing, m
6.9
6.91
6.92
6.93
Easting, m
x 10^5
450
400
350
300
250
200
150
100
μS/cm2


Rank-PCA weighted SOM
2-clusters
Rank-PCA weighted SOM
3-clusters
Rank-PCA weighted SOM
4-clusters
Mouser et al., 2010 classes
M1
M2
M3
M4
M5
M6
M7
M8
M9
M10
M11
M12
M13
M14
M15
M16
M17
M18
M19
M20
M21
M22
Northing, m
Easting, m
x 10^6
x 10^5
1.7135
1.713
1.7125
1.712
1.7115
1.711
1.7105
1.71
6.9
6.91
6.92
6.93
450
400
350
300
250
200
150
100
μS/cm^2

Identifying conditions associated with cyanobacteria blooms in Missisquoi Bay, Lake Champlain, USA, using a modified Self-Organizing Map

Andrea R. Pearce*[1], Mary C. Watzin[2], Gregory K. Druschel[2,3], Lori Stevens[4], Donna M. Rizzo[1]

(1) UVM School of Engineering
(2) UVM Rubsenstein School of Environment and Natural Resources
(3) UVM Geology
(4) UVM Biology

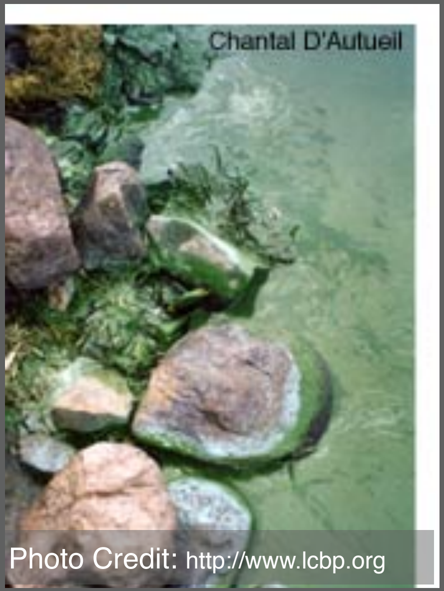


Photo Credit: http://www.lcbp.org

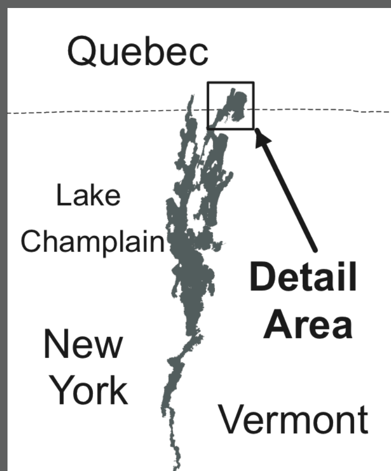
Acknowledgements:

VT EPSCoR Grant NSF EPS #0701410
Kristen Hallock
Alison Pechenik
VT EPSCoR CSYS group

Citation:

A.R. Pearce, M.C. Watzin, G.K. Druschel, L. Stevens, D.M. Rizzo. (In Review), Identifying conditions associtated with cyanobacteria blooms in Missisquoi Bay, Lake Champlain, USA, using a modified Self-Organizing Map. *Limnology and Oceanography*.

The Dataset



- 1 Sampling location
- 2007, 2008, 2009
- Multiple depths (5)
- Physical, Chemical and Biological Measurements
- Concurrent sediment analyses

Missisquoi Bay

◆ Sampling Platform
◆ Long-Term Monitoring Sites

Quickbird image from L. Morrissey

	Parameter	Units	Minimum	Mean	Maximum	Standard Deviation
Nutrients	Total Phosphorus *	μg/L	16.7	49	266	43.5
	Total Nitrogen *	mg/L	0.327	0.579	0.997	0.144
	Dissolved Nitrogen *	mg/L	0.29	0.445	0.997	0.15
	Soluble Reactive Phosphorus *	μg/L	0.5	5.99	33.26	5.46
	Microcystin	μg/L	0.002	2.62	18.5	4.23
Phytoplankton	Bacillariophyceae *	cells/mL	0	306	1694	402
	Chlorophyceae *	cells/mL	23	898	3455	886
	Anabaena *	cells/mL	0	2649	17563	3878
	Aphanizomenon *	cells/mL	0	1027	9919	1960
	Microcystis *	cells/mL	0	25381	319804	58425
Physical Parameters	Temperature *	°C	16.45	21.58	24.44	2.4
	Conductivity	μS/cm	82.15	115.7	127.87	11.09
	Dissolved Oxygen	mg/L	5.98	8.67	11.13	1.25
	PAR (Irradiance)	W/m^2	0.025	212.4	1170	365.6
	Fluorescence	mg/m^3	0.98	7.78	18.86	3.4
	Turbidity	FTU	3.03	10.94	84.01	12.79
Sediment	Dissolved Oxygen *	μM	0	47.4	150	43.8
	Oxic Boundary***	mm	-2.5	-0.66	0.5	0.94
	Mn(II) Redox Boundary***	mm	-26	-1.18	10	7.57

* Also measured as part of the long-term monitoring dataset

** Measured at the sediment water interface by voltametric microelectrode

*** Relative to the sediment-water interface

	Parameter	Units	Minimum	Mean	Maximum	Standard Deviation
Nutrients	Total Phosphorus *	μg/L	16.7	49	266	43.5
	Total Nitrogen *	mg/L	0.327	0.579	0.997	0.144
	Dissolved Nitrogen *	mg/L	0.29	0.445	0.997	0.15
	Soluble Reactive Phosphorus *	μg/L	0.5	5.99	33.26	5.46
	Microcystin	μg/L	0.002	2.62	18.5	4.23
Phytoplankton	Bacillariophyceae *	cells/mL	0	306	1694	402
	Chlorophyceae *	cells/mL	23	898	3455	886
	Anabaena *	cells/mL	0	2649	17563	3878
	Aphanizomenon *	cells/mL	0	1027	9919	1960
	Microcystis *	cells/mL	0	25381	319804	58425
Physical Parameters	Temperature *	°C	16.45	21.58	24.44	2.4
	Conductivity	μS/cm	82.15	115.7	127.87	11.09
	Dissolved Oxygen	mg/L	5.98	8.67	11.13	1.25
	PAR (Irradiance)	W/m^2	0.025	212.4	1170	365.6
	Fluorescence	mg/m^3	0.98	7.78	18.86	3.4
	Turbidity	FTU	3.03	10.94	84.01	12.79
Sediment	Dissolved Oxygen *	μM	0	47.4	150	43.8
	Oxic Boundary***	mm	-2.5	-0.66	0.5	0.94
	Mn(II) Redox Boundary***	mm	-26	-1.18	10	7.57

* Also measured as part of the long-term monitoring dataset

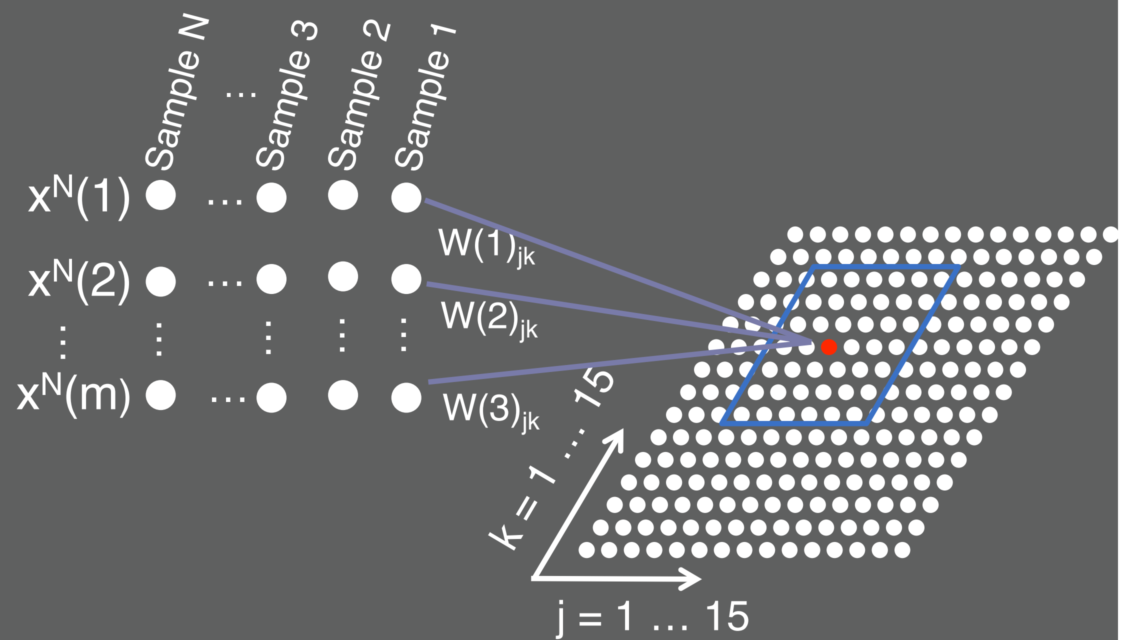
** Measured at the sediment water interface by voltametric microelectrode

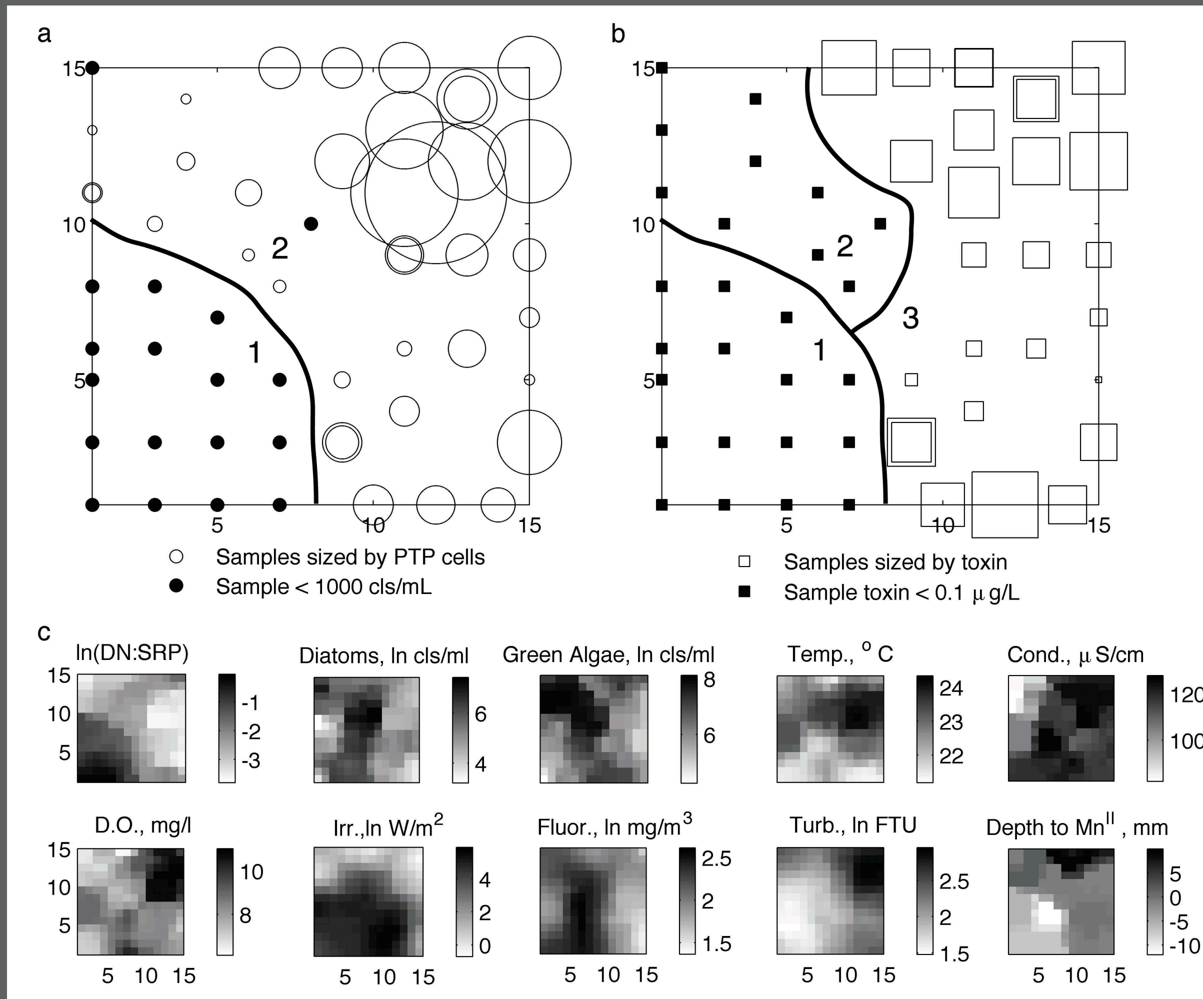
*** Relative to the sediment-water interface

Cluster lake samples based only on phytoplankton cell counts, sediment, and water chemistry.

Input Variables:

1. ln DN:SRP
2. Temp, deg C
3. DO, mg/L
4. Conductivity, μS
5. Fluorescence, mg/m^3
6. Irradiance, W/m^2
7. Turbidity, FTU
8. Green Algae, ln cells/mL
9. Diatoms, ln cells/mL
10. Depth to Dissolved Mn, mm





a
b
1
2
3
Samples sized by PTP cells
Sample < 1000 cls/mL
Samples sized by toxin
Sample toxin < 0.1 μ g/L
c
ln(DN:SRP)
Diatoms, ln cls/ml
Green Algae, ln cls/ml
Temp., ^{o} C
Cond., μ S/cm
D.O., mg/l
Irr.,ln W/m^{2}
Fluor., ln mg/m^{3}
Turb., ln FTU
Depth to Mn^{II} , mm

Conclusions

- Microbial biodiversity in freshwater environments present a unique opportunity for improving characterization and monitoring at sites with multiple data types.
- These new nonparametric methods may be used to describe the spatiotemporal changes associated with microbial community dynamics
- These microbial community profiles in combination with the appropriate computational tools provide substantial value to the physiochemical information traditionally monitored at contaminated water environments

Background on Cyanobacteria in Lake Champlain

- ⌘ Cyanobacteria have always been a part of the phytoplankton community, but recently started dominating in the summer months in several places in the lake.
- ⌘ Microcystis, Anabaena, & Aphanizomenon are the most abundant in Lake Champlain
- ⌘ Cyanobacteria favor nutrient rich water
- ⌘ Suggestions of a minimum threshold of N:P that favors cyanobacteria
  N:P < 29 by weight
- ⌘ Some Cyanobacteria fix N, Microcystis does not
- ⌘ Microcystis can migrate vertically and possibly uptake ammonium from sediment-water interface.

Smith, V. (1983), Low Nitrogen to Phosphorus Ratios Favor Dominance by Blue-Green Algae in Lake Phytoplankton, *Science, 221, 669-671.*